THE FRIEDREICH'S ATAXIA TRIPLET REPEAT SEQUENCE d(GAA)n FORMS A STABLE UNIMOLECULAR STRUCTURE: EVIDENCE FOR A HAIRPIN CONTAINING $A^{anti}.G^{anti}$ BASE PAIRS

By

IANG-SHAN SUEN

Bachelor of Science

Tunghai University

Taichung, Taiwan R.O.China

1992

Submitted to faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
December, 1997

THE FRIEDREICH'S ATAXIA TRIPLET REPEAT SEQUENCE d(GAA)n FORMS A STABLE UNIMOLECULAR STRUCTURE: EVIDENCE FOR A HAIRPIN CONTAINING $A^{anti}.G^{anti}$ BASE PAIRS

Thesis Approved:

Thesis Advisor

Dean of the Graduate College

ACKNOWLEDGMENTS

First of all, I would like to thank God with my whole heart for His guidance and help to finish my MS program. He gave me courage, support and hope to face all the difficulties in my study and research.

With God's blessing, I had an excellent advisor, Dr. Michael Mitas, to enlighten and guide me on my research in Biochemistry field. I wish to express my deep appreciation to him for his intelligent guidance, continuous encouragement and constructive advice. I also want to express my deep gratitude to my committee members, Dr. Ulrich Melcher and Dr. Steven Graham for their advice and critical suggestions that are really helpful to my training and study.

I would like to thank the faculty of the Department of Biochemistry and Molecular Biology for their teaching and friendship. They played an important role in building up my knowledge in this field. I also thank my lab colleagues, Brian McEwen and Mellisa Christy for their help and assistance through my research.

Finally I want to express my deepest gratitude to my parents and sister. Their endless love, understanding, support and prayer were the sweetest part of my research. I dedicate my dissertation to them with my love.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

LIST OF ABBREVIATIONS

A	adenine
C	cytosine
DEPC	diethylpryrocarbonate
DMS	dimethyl sulfate
ds	double-stranded
EMMP	electrophoretic mobility melting profiles
G	guanine
HA	hydroxyl amine
ss	single-stranded
T	thymine
TREDs	Triplet Repeat Expansion Diseases
T_m	melting temperature
UTR	Untranslated Region
W-C	Watson-Crick

CHAPTER I

Introduction

Triplet repeat DNA expansion: a new mechanism that underlies many human inherited diseases

Triplet repeat expansion diseases (TREDs) are regarded as a unique family among inherited genetic disorders (Ashley & Warren, 1995). Typically, they are characterized by expansion of genetically unstable d(CTG•CAG), d(CCG•CGG) or d(GAA•TTC) triplet repeat sequences located within specific genes. The number of triplet repeats in disease-associated genes tends to increase in germline transmission from parent to offspring. The amplification continues in offspring of affected individuals, resulting in progressively increased severity of the disease and/or an earlier age of onset, phenomena clinically referred to as "anticipation". Members of the TRED family include Huntington's disease, fragile X syndrome and Friedreich's ataxia [for a review of the genetic and clinical features of TREDs, see Ashley & Warren (1995); Sutherland & Richards (1995); Timchenko & Caskey (1996)]. Usually, the larger the repeat number, the higher the

probability of expansion and the earlier age of onset and increased severity of the disease. Due to the high frequency with which the trinucleotide repeats associated with these diseases undergo further expansion in offspring of affected individuals, the mutation(s) responsible for this phenomenon has been described as "dynamic" (Sutherland and Richards, 1995).

Individuals with a fragile site at Xq27.3 develop mental retardation. Hence, this disease was coined "fragile X syndrome". In 1991, three groups of researchers independently reported a large amplifcation of $d(CGG)_n$ trinucleotide repeats as the cause of fragile X syndrome (Bell et al., 1991; Vincent et al., 1991; Pierreti et al., 1991). Since then, at least ten other human inherited diseases have also been found associated with triplet repeat DNA expansion. Most of these diseases are neurological disorders caused by a $d(CAG)_n$ triplet expanded in the translated regions of their respective genes.

Of all possible triplet repeat sequences, only three have been found to be associated with human disease; these are CGG/CCG, CAG/CTG, and GAA/TTC. Most trinucleotide repeats expanded in human genetic disorders occur in exons or 5' and 3' untranslated regions (UTRs). Expansion of a simple tandem repeat within an intron associated with a human genetic disorder was only found with GAA (Campuzano et al., 1996).

$(CGG)_n$ and $(CAG)_n$ triplet repeats are found in close proximity to numerous gene loci (Riggins et al., 1992; Li et

al., 1993). The CGG repeats are responsible for one group of fragile sites on 5′ untranslated region of *FMR1* gene of the X chromosomes and can expand to very high copy numbers, in excess of 1000 copies. Expansion of the d(CGG) repeat is accompanied by methylation of the *FMR1* promoter and of the amplified trinucleotide tract (Bell et al., 1991; Vincent et al 1991; Pierreti et al., 1991). Subsequent to d(CGG) expansion and hypermethylation, the *FMR1* gene becomes transcriptionally silent (Pierreti et al 1991; Hansen et al., 1992; Sutcliffe et al., 1992). The CAG repeats are involved in a number of neurological disorders. They can also expand to high copy numbers when in the untranslated region of a gene (as in myotonic dystrophy) but are mostly in coding regions where their copy number is usually less than 100. Expansion of $(CAG)_n$ in the corresponding gene usually does not cause inactivity of the gene, but results in abberant translation products (Sutherland and Richards, 1995; Warren, 1996).

Characteristics of Friedreich's ataxia

Friedreich's ataxia (Friedreich, 1863), an autosomal recessive disease, was recently shown to be associated with expansion of an intronic GAA triplet repeat located in the frataxin gene (Campuzano et al., 1996).

Friedreich's ataxia is the most common early-onset hereditary ataxia, with an estimated prevalence of 1 in 50,000 and a deduced carrier ranging from 1 in 110 to 1 in 70 (Harding et al., 1982; Filla et al., 1992). Friedreich's ataxia is characterized by progressive ataxia with onset before adolescence, incoordination of limb movements, dysarthria, nystagmus, diminished or absent tendon reflexes, Babinski sign, impairment of position and vibratory senses, skeletal deformities, cardiomyopathy, and diabetes (Geoffroy et al., 1976). Friedreich's ataxia has a higher clinical variability than expected in an autosomal recessive disorder, in particular concerning age at onset and presence of cardiomyopathy and diabetes (Filla et al., 1992).

Campuzano et al. (1996) isolated the defective gene X25 causing Friedreich's ataxia. They found X25 encodes a 210-amino acid protein with unknown function called "frataxin". They also demonstrated reduction of frataxin mRNA levels in Friedreich's ataxia patients. An expanded unstable $d(GAA)_n$ repeat is located in the X25 first intron on Friedreich's ataxia chromosomes. In normal individuals the number of repeats varies from 7 to 22, whereas Friedreich's ataxia patients carry 200 to above 900 GAA repeats. In addition, three rare point mutations have been described, especially an amino acid change (isoleucine to phenylalanine at point 154, I154F) found in X25 exon 4 in southern Italian patients (Campuzano et al., 1996; Filla, et al, 1996).

Recently, Carvajal et al. reported that the STM7 gene rather than X25 is the candidate for Friedreich's ataxia (Carvajal et al., 1996). They demonstrated that X25 comprises part of the STM7 gene, contributing at least four splice variants. The expanded intronic GAA triplet repeat in affected individuals located downstream of X25 exon 1, which now was reassigned as exon 18 of the STM7 gene. In addition, their data did not show evidence for a reduction in the level the mature mRNA incorporating exons spanning the repeat motif in affected individuals as Campuzano et al. demonstrated (Carvajal et al., 1996). Further investigation of the expressed STM7 recombinant protein corresponding to the reported 2.7 kb transcript (STM7.1) demonstrated a novel phosphatidylinositol-4-phosphate 5-kinase activity (Carvajal et al., 1996), implicating the pathogenesis of the disease due to the defect in the phosphoinositide regulated signaling pathways, possibly affecting vesicular trafficking or synaptic transmission (Carvajal et al., 1995; De Camilli, et al., 1996).

Proposed expansion mechanism: hairpin formation during DNA replication

Biophysical and biochemical studies have shown that GC-rich triplet repeat sequences [$d(CTG)_n$, $d(CAG)_n$, $d(CCG)_n$, $d(CGG)_n$, $d(GTC)_n$, and $d(GAC)_n$], which undergo expansion in human genes (Ashley & Warren, 1995) or

duplication/recombination in bacterial systems (Ohshima et al., 1996), form stable hairpins (Chen et al., 1995; Gacy et al., 1995; Gao et al., 1995; Mariappan et al., 1996a; Mariappan et al., 1996b; Mitas et al., 1995a; Mitas et al., 1995b; Petruska et al., 1996; Smith et al., 1995; Yu et al., 1997a; Yu et al., 1995a; Yu et al., 1995b). The hairpin formed by $d(CGG)_n$ has the additional capability of folding into a tetraplex structure (Usdin & Woodford, 1995). These findings have led to the conclusion that hairpin (or hairpin-like) structures form at the replication fork (Chen et al., 1995; Gacy et al., 1995; Kang et al., 1996; Mariappan et al., 1996b; Mitas et al., 1995b; Pearson & Sinden, 1996) [presumably at the lagging daughter strand (Kang et al., 1995)] and play a major role in expansion events.

Since $d(GAA)_n$ repeats, whose expansion is associated with Friedreich's ataxia(Campuzano et al., 1996), seem apparently incapable of forming hairpin structures similar to those containing CG-rich triplet repeats, it was argued (Warren, 1996) that DNA structures play only a minor (if any) role in expansion events and further, that all triplet repeat sequences [including those containing no secondary structure (Mitas, 1995)] might be susceptible to expansion. Others have argued that there might be two separate mechanisms of expansion (Timchenko & Caskey, 1996), the first of which involves hairpin structures. The second mechanism (McMurray et al., 1997) might involve intramolecular triplexes (Mariappan et al., 1997; Ohshima et al., 1996a) containing

the Friedreich's ataxia $d(GAA)_n$ repeats, opening up the possibility that other triplex-forming minisatellite sequences might be susceptible to expansion. However, in contrast to these possibilities, a more parsimonious explanation is that expansion of a minisatellite sequence requires the formation of a unimolecular hairpin (Mitas, 1997), and the one adopted by Friedreich's ataxia $d(GAA)_n$ uniquely lacks Watson-Crick base pairs.

Since DNA structure lies at the heart of understanding mechanisms of minisatellite expansion, it is important to determine whether $d(GAA)_n$ is capable of adopting a hairpin conformation. Such a hairpin might contain one of three types of A•G base pairs that have been observed in crystal structures of duplex DNA (Kennard & Hunter, 1989). Each of these A•G base pairs [A^{syn}•G^{anti} (Brown et al., 1986; Hunter et al., 1986), A^{anti}•G^{anti} (Brown et al., 1989), and A^{+anti}•G^{syn} (Prive et al., 1987)] contains two H-bonds, one of which is adenine N6-guanine O6 (Fig. 1). The largest distortion of the sugar-phosphate backbone is observed for the A^{anti}•G^{anti} pairing where the C1'-C1' separation increases from canonical B-DNA by almost 2Å (Kennard & Hunter, 1989).

To resolve the issue of whether Friedreich's ataxia $d(GAA)_n$ repeats adopt a hairpin conformation, we investigated the structure of $d(GAA)_n$ by electrophoretic, chemical modification, and P1 nuclease digestion techniques.

CHAPTER II

Materials and Methods

Oligonucleotides

Oligonucleotides were synthesized by the OSU Recombinant DNA/Protein Resource Facility on an Applied Biosystems 381A oligonucleotide synthesizer (Foster City, CA) with the trityl group on and purified with oligonucleotide purification cartridges (Cruachem, Glasgow, UK). Sequences of oligonucleotides were:

Oligo (***1***) : AGCTTTGATCA$(GAA)_{15}$G

Oligo (***1'***): GATC$(CTT)_{15}$CTGATCAA

Oligo (***2***) : AGCTTTGATCA(GAA)(GAT)$(GAA)_{14}$GGATC

Oligo (***2'***): GATCC$(TTC)_{14}$(ATC)(TTC)TGATCAAAGCT

Oligo (***3***) : AGCTTGTTAACC$(GAA)_{15}$G

Oligo (***3'***): GATCC$(TTC)_{15}$GGTTAACA.

Plasmid DNA Preparation

Five μg of each synthetic oligonucleotide was phosphorylated with ATP and T4 polynucleotide kinase (Boehringer Mannheim,

Indianapolis, IN) in buffer containing 1 mM DTT, 1 mM ATP, 50 mM Tris-HCl (pH 7.6), 10 mM $MgCl_2$, 0.1 mM spermidine, 0.1 mM EDTA and 20 units of T4 polynucleotide kinase (New England Biolabs, Beverly, MA) in a final volume of 100 µl. Complementary pairs were combined, heated to 100°C, and allowed to cool to room temperature over a period of at least 4 hours and ligated into the *Bam*HI and *Hin*dIII sites of pBluescript SK^+ (Stratagene, La Jolla CA) as previously described (Mitas et al., 1995 b). Plasmid DNA was transformed into competent XL-1Blue cells (Stratagene) and purified by cesium chloride centrifugation (Sambrook et al., 1989).

DNA Sequence analysis

DNA sequence analysis was performed by OSU Recombinant DNA/Protein resource facility on an ABI 373 automated sequencer with the use of the M13 universal primer. The insert of one plasmid arising from ligation of double-stranded oligonucleotide (***1***)•(***1'***) contained no mutations and was named pGAA1. The sequence of another insert was 5' AAGCTTTGATCA(GAA)(GAT)$(GAA)_{14}$GGATCC. The plasmid bearing this sequence was named pGAA2. The insert of a plasmid arising from ligation of double-stranded oligonucleotide (***3***)•(***3'***) contained no mutations and was named pGAA3.

Plasmid Labeling

Labelings of oligonucleotides liberated from pGAA1-3 were performed with the use of α-^{32}P-dCTP, α-^{32}P-dATP (ICN, Irvine, CA) and DNA polymerase I Klenow fragment (New England Biolabs, Beverly, MA) as previously described (Yu et al, 1995). For labeling of the GAA-containing strands, pGAA1 and pGAA2 were first cleaved with *Bam*HI, labeled with ^{32}P, and then cleaved with *Hin*dIII. The sequences of the labeled single strands containing GAA repeats derived from pGAA1 and pGAA2 are referred to as oligos ***1*** and ***2***, respectively, and are listed in table 1. Labeling of the GAA-containing strand derived from pGAA3 was performed in a similar manner, except that *Hpa*I was used in place of *Hin*dIII. This labeled strand is referred to as oligo ***3***. For labeling of the TTC-containing strands, pGAA1 and pGAA2 were first cleaved with *Hin*dIII, labeled with ^{32}P as described above, and then cleaved with *Bam*HI. The sequences of the labeled single strands containing TTC repeats derived from pGAA1 and pGAA2 are referred to as oligos ***1'*** and ***2'***, respectively (table 1).

P1 nuclease digestion

Labeled oligonucleotides were separated from labeled vector by electrophoresis in a 0.8% agarose gel. Oligonucleotides containing triplet repeats were excised from gels and purified with glass beads (Mermaid Kit, Bio101, La Jolla, CA). 3' end-labeled oligos ***1, 2, 2'*** and ***3*** (4 x 10^3

dpm; 0.7 fmoles) were added to synthetic oligos (1.4 pmoles) (***1***), (***1'***) and (***3***) and placed in a 70 °C water bath for 3 min. and placed on ice for 5 min. P1 nuclease digestions were performed in 12 µl of solution (pH 6.5, 7.5 and 8.5, respectively), containing 50 mM sodium acetate, 2 mM $ZnSO_4$, 5% glycerol and various amount of P1 nuclease at various temperatures essentially according to the method of Wohlrab (Wohlrab, 1992) as previously described (Mitas et al., 1995b). Various amounts of NaCl were added to differrent reactions to get higher Na^+ concentrations. The reactions were stopped by the addition of 1.3 µl of 0.5 M EDTA solution.

Chemical modification with hydroxylamine

Oligo ***2'*** was labeled as described above for the P1 nuclease analysis. Hydroxylamine (HA) reactions were performed essentially according to the method of Rubin & Schmid (Rubin & Schmid, 1980). Briefly, HA was freshly prepared by titrating 4 M hydroxylamine hydrochloride (Aldrich) to the required pH (6.5-8.5) with diethylamine (Sigma). Seven µl of labeled oligo ***2'*** (2.0×10^4 dpm, ~0.7 fmoles) and 1.4 pmoles of synthetic oligo (***3'***) was heated to 70 °C for 3 minutes and immediately placed on ice. Various amounts of HA (6-46 µl) were added to the DNA and incubated for 25 min at 37 °C. NaCl was added to yield a final concentration of 50 mM. The reaction was stopped by the

addition of a 0.25 ml stop solution [0.3 M sodium acetate (pH 5.2), 0.1 mM EDTA, 25 µg/ml tRNA] and 0.75 ml cold ethanol. The precipitated DNA was resuspended in 100 µl 1 M piperidine and heated for 30 min at 92 °C to generate strand breaks. After removal of piperidine *in vacuo* (x2), the dried DNA was resuspended in formamide loading buffer [80% formamide, 10mM NaOH, 1mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue], and loaded on a DNA sequencing gel containing 20% polyacrylamide and 8 M urea.

Chemical modification with diethylpyrocarbonate

Oligo ***2*** and ***3*** were labeled and prepared as described above for the P1 nuclease analysis. Diethyl pyrocarbonate reactions were performed essentially according to the method of McCarthy et al. (McCarthy et al., 1990). Modification reactions were performed for different periods of time with various amounts of DEPC (2.7-7.1 µl) in buffer containing 50mM Na^+ cacodylate (pH 7.5), 2 mM EDTA. The reaction was stopped by the addition of 50 µl stop solution [1.5 M sodium acetate(pH 5.2) ,200 µg/ml tRNA] and 0.75 ml cold ethanol. The subsequent procedures were performed as described above for the hydroxylamine reaction.

Chemical modification with dimethyl sulfate

Oligo ***2*** and ***3*** were labeled and prepared as described above for the P1 nuclease analysis. Dimethyl sulfate (DMS) reactions were performed at 21 mM with the DNAs at various temperatures essentially according to the method of Maxam and Gilbert (Maxam & Gilbert, 1980). Briefly, modification reactions were performed for different periods of time in buffer containing 50mM Na^+ cacodylate (pH 7.5), 2 mM EDTA. The reaction was stopped by the addition of 15 µl stop solution [1.5 M sodium acetate (pH 7.0), 1.0 M mercaptoethanol, 100 µg/ml tRNA] and 0.75 ml cold ethanol. The subsequent procedures were performed as described above for the hydroxylamine reaction.

CHAPTER III

Results and discusion

To analyze DNA containing d(GAA) and d(TTC) repeats, double-stranded (ds) oligonucleotides containing $d(GAA)_{15}$ were cloned into plasmids as described in materials and methods. Oligonucleotides liberated from plasmids were utilized for studies since they were unequivocally full-length. For clarity, the triplet repeats within the ss oligonucleotides are indicated in the subsequent text by roman numerals. The sequences of the cloned oligonucleotides are listed in table 1. The GAA repeat-containing strands are designated as oligos ***1***, ***2***, and ***3***. The TTC repeat-containing complementary strands (of oligos ***1***, ***2*** and ***3***) are designated as oligo ***1'***, ***2'*** and ***3'***, respectively.

P1 nuclease analysis and chemical modification indicated that oligo *2'* contained no preferred secondary structure

Preliminary studies conducted in the Mitas lab by B. McEwen and M. Christy showed that oligo ***2'*** contained no preferred secondary structure (data not shown).

To further investigate the possibility that oligo ***2'*** contained no preferred secondary structure, chemical modification experiments were performed with hydroxylamine (HA), which reacts with cytosines that are not base-paired (Johnston, 1992, Johnston & Rich, 1985, Singer & Grunberger, 1983). Also, in order to detect regions of oligo ***2'*** that were single-stranded, it was incubated with P1, an endonuclease that cleaves random bases in ssDNA (Wohlrab, 1992).

Consistent with the electrophoretic data that oligo ***2'*** lacked a preferred secondary structure, HA modified all cytosines at 37 °C, 50 mM Na^+, pH 7.5 (Fig. 2). Furthermore, P1 nuclease was able to cleave the phosphodiester bonds of oligo ***2'*** (Fig. 2). However, sequence preferences were detected: the TpT phosphodiester bonds were cleaved less efficiently than CpT or TpC bonds.

Oligo *3* adopts an asymmetric hairpin conformation at 5 °C

The electrophoretic analysis performed by B. McEwen and M. Christy suggested that oligo ***3*** contained secondary structure (data not shown). Five potential structures of oligo ***3*** are

shown in figure 3, three of which (structures ***A***, ***B***, ***D***) involve folding of the d(GAA) repeat into a hairpin. To investigate the possibility that oligo ***3*** adopted any of the five potential conformations, it was incubated with diethylpyrocarbonate (DEPC), which preferentially alkylates the N7 of adenine residues that are not involved in stacking interactions (Krol & Carbon, 1989). At 5 °C, the adenines in triplet VIII were extensively modified when incubated with DEPC for 30 or 60 min, providing evidence for a hairpin conformation. This result suggests that oligo ***3*** does not form structures ***C*** or ***E*** (Fig. 3), which are arranged in linear arrays. The adenines in triplet VII, as well as the adenines at nucleotide positions 6 and 9 were also highly susceptible to modification. Comparison of the products generated at 60 and 120 min (5 °C) suggested that extensive modification of the adenines in triplets VII and VIII at 60 min appeared to be due to *hypersensitivity* of the respective bases, rather than protection of the remaining bases. The reason for lack of modification of triplets VII and VIII at 120 min is not clear, but suggests that minor experimental perturbations may influence detection of the hairpin structure.

The degree to which the remaining adenines reacted with DEPC was dependent upon their location in the hairpin. The 5' adenine in a given 5'-GAA-3' triplet was more reactive compared to the 3' adenine (Fig. 4).

At 15-45 °C, the reactivity of the adenines in the central portion of the triplet repeat decreased, providing evidence

that either the structure of oligo ***3*** did not contain a loop, or alternatively, that the adenines in the loop were not hypersensitive to DEPC. Identical results were obtained when oligo ***3*** was incubated with DEPC for times shorter or longer than those depicted in figure 4 (data not shown). However, similar to the results obtained at 5 °C, the 5' adenines within triplets were more reactive compared to the neighboring 3' adenines. This result provides evidence the structure of oligo ***3*** formed at 25-45 °C is similar to the hairpin formed at 5 °C.

P1 nuclease cleaves the phosphodiester bonds in the central region of the triplet repeat

To investigate further the structure of oligo ***3***, it was incubated with single-strand-specific P1 nuclease in various reaction coditions.

At 5 °C (Fig. 5) and 15 °C (Fig. 6), P1 nuclease cleavage sites were observed in the middle of the triplet repeat region (at the G26pA27, A27pA28 and A28pG29 phosphodiester bonds), indicating that oligo *3* folded into a hairpin structure at these temperatures. However, the amounts of cleavages of these three phosphodiester bonds appeared minor compared to the 3'-end of the sequence, providing evidence that the nucleotides in the loop were involved in extensive stacking interactions. Previous experiments conducted with

hairpin structures containing triplet repeats have shown that the phosphodiester most susceptible to cleavage by P1 nuclease is the one located 3' to the last base pair (Mitas et al., 1995a, Yu et al., 1997a, Yu et al., 1995a, Yu et al., 1995b). Therefore, we anticipated that if oligo ***3*** adopted structure ***A***, the major site of P1 nuclease cleavage would be the A49pG50 phosphodiester bond. In contrast, the G50pG51 phosphodiester would be preferentially cleaved in structures ***B*** or ***D***. As shown in Fig. 5, two major sites of P1 nuclease digestion were detected 3' to the triplet repeat region, both of which involved purine dinucleotides (G50pG51 and G51pA52). This result is consistent with structures ***B*** and ***D***. Also, the observation that P1 nuclease was capable of cleaving the purine dinucleotides G50pG51 and G51pA52 rules out the possibility that lack of cleavage observed in the triplet repeat region was due to purine content.

At 25 °C, P1 nuclease failed to cleave the phosphodiester linkages in the triplet repeat region of oligo ***3*** (with the exception of minor cleavage in triplet XV) (Fig. 5). Similarly, P1 nuclease also failed to cleave the triplet repeat region of oligo ***3*** at 37 °C, but cleaved all phosphodiester linkages 3' to the repeat (Fig. 7).

Cleavages of the three central phosphodiesters were also observed at 5 °C at pH 6.5 or 8.5 (Fig. 8). The result at the higher pH provides evidence that the hairpin structure does not require protonation of the adenine residues, which is necessary for a hairpin containing $A^{+anti} \cdot G^{syn}$ base pairs

(Prive et al., 1987) or for a helical single-stranded structure containing alternating As and Gs (Dolinnaya & Fresco, 1992, Mukerji et al., 1995, Shiber et al., 1995).

At 5 °C, 150 mM Na^+, pH 7.5, the sites of P1 nuclease cleavage were similar to those generated in 50 mM Na^+, with the exception that the cleavage at the A28pG29 phosphodiester was more pronounced (Fig. 9). For oligo ***2***, the cleavage patterns were also essentially the same at 37 °C with various amount of Na^+ (Fig. 10). The results of P1 nuclease digestion studies provide evidence that the structure of $d(GAA)_n$ is relatively stable in low and moderate salt, neutral pH, but is sensitive to heat. These properties are consistent with a nucleic acid structure that is primarily stabilized by H-bonds.

The hairpin structure of oligo *3* does not contain $G^{syn}\cdot G^{anti}$ base pairs

The pattern of P1 nuclease cleavage at 5 °C is most consistent with hairpin structures ***B*** and ***D***. The ***D*** hairpin contains $G^{syn}\cdot G^{anti}$ base pairs, whereas the ***B*** hairpin contains $A^{anti}\cdot G^{anti}$ or $A^{syn}\cdot G^{anti}$ base pairs ($A^{+anti}\cdot G^{syn}$ pairs can be ruled out for the reasons described above). Due to the presence of $G^{syn}\cdot G^{anti}$ pairs, a guanine in the stem region of a ***D*** hairpin is expected to be at least partially resistant to modification by dimethyl sulfate (DMS). In contrast, the guanines in the stem region of a ***D*** hairpin (all of which are

predicted to be in the *anti* conformation) will not be resistant to modification by DMS. Therefore, the use of DMS can theoretically differentiate between hairpins in ***B*** and ***D*** alignments. In 50 mM Na^+ at 25 °C (Fig. 5) or 5 °C (Fig. 9), all guanines in oligo ***3*** uniformly reacted with DMS, providing evidence that this sequence lacked $G^{syn}{\bullet}G^{anti}$ pairs. Similar results were also observed in oligo 2 (Fig. 11). Hence, we conclude that oligo ***3*** adopts a hairpin in the ***B*** alignment.

Discussion

Previous studies have shown that $d(GA)_n$ adopts a variety of structural conformations, including a hairpin duplex (Casasnovas et al., 1993, Huertas et al., 1993, Shiber et al., 1996), a two-hairpin tetraplex (Shiber et al., 1996, Mukerji et al., 1996), parallel-stranded duplexes (Rippe et al., 1992, Robinson et al., 1994, Dolinnaya et al., 1997) and a novel type of single-standed helix (Dolinnaya et al., 1993, Dolinnaya & Fresco, 1992, Shiber et al., 1995, Mukerji et al., 1995). The structure adopted by $d(GA)_n$ is dependent upon DNA concentration, nucleotide length, temperature, monovalent cation concentration and pH. At low strand concentration ($6x10^{-4}$ mM), neutral pH, low temperature (5-15 °C) and physiologic salt concentration (150 mM Na^+), $d(GA)_{10}$ adopts a hairpin conformation as evidenced by relatively rapid electrophoretic mobility and cleavage in the central portion of the sequence by S1 nuclease (Shiber et al., 1996). The base pairing arrangement of the hairpin structure is not known. Increasing the length of $d(GA)_n$ from n=10 to n=20 results in formation of a parallel-stranded duplex structure that at neutral pH and moderate salt, contains A•A and G•G base pairs (Dolinnaya et al., 1997). The T_m of the $d(GA)_{20}$•$d(GA)_{20}$ duplex in 0.21 M Na^+ (pH 7.5) is 43.3 °C (Dolinnaya et al., 1997).

Similar to the results obtained by Fresco and colleagues (Shiber et al., 1996), we have provided evidence that at 5 °C, pH 7.5, 50-150 mM NaCl, $d(GAA)_{15}$ (contained within oligo ***3***), adopts a hairpin conformation. Data supporting a hairpin conformation were increased electrophoretic mobility of oligo ***3*** (data not shown), P1 nuclease digestion of three phosphodiester linkages in the presumed hairpin loop region (Fig. 5), and hypersensitivity of the adenines in triplets VII and VIII to modification by DEPC (Fig. 4). An estimate of the T_m of oligo ***3*** from electrophoretic data was 50 °C (data not shown), providing evidence that the structure adopted by this sequence is more stable than either the $d(GA)_{10}$ hairpin or the $d(GA)_{20}$•$d(GA)_{20}$ parallel-stranded duplex.

In contrast to the adenine residues, the guanines in the loop region were not hypersensitive to chemical modification at 5 °C (Fig. 9). Thus, we are fairly confident that the $d(GAA)_{15}$ hairpin does not contain G^{syn}•G^{anti} pairs. On the basis of this result and the P1 nuclease cleavage pattern at the 3'-end of the oligo ***3*** (Fig. 5) and oligo ***2*** (Fig. 10), we have tentatively assigned a base-pairing arrangement of the antiparallel strands in the $d(GAA)_{15}$ hairpin such that a 5'-GAA-3' repeat on the 5'-side is paired to a 5'-AGA-3' repeat on the 3' side (ie., a hairpin containing A•G and A•A pairs in a 2:1 ratio).

The d(GAA)$_n$ hairpin may contain A^{anti}•G^{anti} base pairs

In the 5'-half of the hairpin structure, the 3' adenines of a 5'-GAA-3' repeat (which are predicted to be part of an A•A pair) were relatively resistant to modification by DEPC (Fig. 4). This result suggests that the putative A•A pairs within the hairpin stem were well stacked in the helix and that one of the two N7 atoms within a pair might be involved in H-bond formation (ie., the one on the 5'-side of the hairpin). In contrast, the 5' adenines (which are predicted to be part of an A•G pair) of a 5'-GAA-3' repeat were highly susceptible to DEPC modification. Since this result is not consistent with an A•G base pair in which the adenine N7 is in the *syn* conformation, we suspect that the conformation adopted by the pair is A^{anti}•G^{anti}.

The structure formed by oligo ***3*** at physiologic temperature is not known. The EMMP data obtained by B. McEwen and M. Christy indicated that oligo ***3*** did not undergo a structural transition between 3 and 41 °C (data not shown), providing evidence that only one conformation was adopted over this temperature range. In addition, the patterns of P1 nuclease cleavages at the 3' end of oligo ***3*** at 5 °C (Fig. 5) and 37 °C (Fig. 7) were essentially identical and consistent with a hairpin conformation. However, at physiologic temperature, we could not obtain evidence by P1 nuclease digestion analysis or by chemical modification that a loop was present in the

middle of $d(GAA)_{15}$. A possible explanation for this result is that the bases in the loop region form a well stacked structure at 37 °C. In support of this possibility, studies have shown that GdAGA (Orita et al., 1996), d(GAAA) (Heus & Pardi, 1991) or d(GNA) (Yoshizawa et al., 1997) loops contained in hairpin structures are stabilized by extensive stacking interactions. Another possible explanation is that at 37 °C, oligo ***3*** forms a stable intramolecular structure that is not a hairpin.

Hairpin structure of $d(GAA)_n$: implications regarding potential mechanisms of minisatellite expansion

In this report we have provided evidence that similar to other triplet repeats associated with human disease, $d(GAA)_n$ is capable of adopting a hairpin structure *in vitro*. This finding has several important implications regarding potential mechanisms of minisatellite expansion. First, it suggests that hairpin or hairpin-like structures are required for expansion events. Hairpins may play an indirect role by preventing translocation of DNA helicases, which maintain their polarity at the replication fork by binding to both ss- and dsDNA [reviewed in Lohman & Bjornson, (1996)]. It is well documented that blocks to DNA replication, which are frequently caused by palindromic sequences that form hairpin or cruciform structures, lead to deletions, translocations,

insertions, duplications or rearrangements of DNA [reviewed by Bierne & Michel (1994), but see also Gordenin et al., (1993)]. Alternatively (but not necessarily mutually exclusive), hairpin structures could play a direct role in the expansion process by facilitating DNA slippage during lagging strand synthesis. The 3'- and 5'-ends of the resultant hairpin would be in close proximity to one another, an arrangement that is mirrored in intramolecular tetraplex, but not triplex, structures. This arrangement would be amenable to subsequent ligation of the 5'-end of the hairpin to a newly synthesized Okazaki fragment, resulting in the creation of complementary strands of DNA of unequal length. Second, hairpins, hairpin-like structures, or the DNA strands complementary to these structures may be targeted by structure-specific repair enzymes, similar to those involved in hairpin-mediated V(D)J cleavage and recombination (Hiom & Gellert, 1997, McBlane et al., 1995). The repair of other DNA structures such as triplexes, which have been convincingly shown to block DNA replication *in vitro* (Krasilnikov et al., 1997) and *in vivo*, (Baran et al., 1987, Rao, 1994) would likely require a different set of repair enzymes.

FEN1, a mammalian homologue of RAD27 endonuclease in yeast, was found able to remove a 5′ flap generated by displacement synthesis while DNA polymerase encounters the 5′ end of a downstream Okazaki fragment (Bambara et al., 1997, Lieber M. R., 1997). When a 5′ flap generated, the FEN1

endonuclease is loaded at the 5′ end of the flap similar to a bead with a fine bore stringing up a thread and slides down to the base of the flap, then remove the flap. In vitro studies showed that FEN1 endonuclease can only act on single-stranded DNA ,and a partially double-stranded flap would block FEN1 fuction (Bambara et al., 1997, Lieber M. R., 1997). Recently, a model proposed by Gordenin et al (1997) illstrated that the formation of FEN1-resistant secondary structures, like a hairpin or a triplex, at a 5′ flap generated by displacement synthesis during DNA replication of the lagging strand would result in expansion. Hairpin structures and/or other kind of secondary structures adopted by d(GAA)15 at 5oC and/or higher temperatures implicate the possible mechanism of triplet expansion involved in this model.

REFERENCE

Ashley, J. C. T., & Warren, S. T. (1995) *Annu. Rev. Genetics 29*, 703-728.

Bambara, R. A., Murante, R. S., & Henricksen, L. A. (1997) J. Biol. Chem. 272, 4647-4650.

Baran, N., Lapidot, A., & Manor, H. (1987) *Mol. Cell. Biol. 7*, 2636-2640.

Bell, M. V., Hirst, M. C., Nakahori, Y., MacKinnon, R. N., Roche, A., Flint, T.J., Jacobs, P. A., Tommerup, N., Tranebjaerg, L., Froster-Iskenius, U., Kerr, B., Turner, G., Lindenbaum, R. H., Winter, R., Pembrey, M., Thibodeau, S., & Davies, K. E. (1991) Cell 64, 861-866.

Bierne, H., & Michel, B. (1994) *Mol. Micro. 13*, 17-23.

Brown, T., Hunter, W. N., Kneale, G., & Kennard, O. (1986) *Proc. Natl. Acad. Sci. 83*, 2402-2406.

Brown, T., Leonard, G. A., Booth, E. D., & Chambers, J. (1989) *J. Mol. Biol. 207*, 445-457.

Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Cañizares, J., Koutnikova, H., Bidichandani, S. I., Gellera, C., Brice, A., Troullas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P. I., Di Donato, S., Mandel, J.-L., Cocozza, S., Koenig, M., & Pandolfo, M. (1996) *Science 271*, 1423-1427.

Carvajal, J. J., Pook, M. A., Santos, M. D., Doundey, K., Hillermann, R., Minogue, S., Williamson, R., Hsuan, J. J., & Chamberlain, S. (1996) *Nat. Genet. 14*, 157-162.

Carvajal, J. J. (1995) *Hum. Molec. Genet. 4*, 1411-1419.

Casanovas, J. M., Huertas, D., Ortiz-Lombardia, M., Kypr, J., & Azorin, F. (1993) *J. Mol. Biol. 233*, 671-681.

Chen, X., Mariappan, S. V. S., Catasti, P., Ratliff, R., Moyzis, R. K., Laayoun, A., Smith, S. S., Bradbury, E. M., & Gupta, G. (1995) *Proc. Natl. Acad. Sci. USA 92*, 5199-5203.

De Camilli, P., Emr, S. D., McPherson, P. S. ,& Novick, P. (1996) *Science 271*, 1533-1539

Dolinnaya, N. G., Braswell, E. H., Fossella, J. A., Klump, H., & Fresco, J. R. (1993) *Biochemistry 32*, 10263-10270.

Dolinnaya, N. G., & Fresco, J. R. (1992) *Proc. Natl. Acad. U.S.A. 89*, 9242-9246.

Dolinnaya, N. G., Ulku, A., & Fresco, J. R. (1997) *Nucleic Acids Res. 25*, 1100-1107.

Filla, A., De Michele, G., Marconi, R., Bucci, L., Carillo, C., Castellano, A. E., & Iorio, L. (1992) *J. Neurol. 239*, 351-353.

Filla, A., De Michele, G., Cavalcanti, F., Pianese, L., Monticelli, A., Campanella, G., & Cocozza, S. (1996) *Am. J. Hum. Genet. 59*, 554-560.

Friedreich, N., (1863) *Virchows Arch. Pathol. Anat. 26*, 391.

Gacy, A. M., Goeliner, G., Juranic, N., Macura, S., & McMurray, C. T. (1995) *Cell 81*, 533-540.

Gao, X., Huang, X., Smith, K. G., Zheng, M., & Liu, H. (1995) *J. Am. Chem. Soc. 114*, 8883-8884.

Geoffroy, G., Barbeau, A., Breton, G., Lemieux, B., Aube, M., Leger, C., & Bouchard, J. B. (1976) *J. Neurol. Sci. 3*, 279-286

Gordenin, D. A., Lobachev, K. S., Degtyareva, N. P., Malkova, A. L., Perkins, E., & Resnick, M. A. (1993) *Mol. Cell. Biol. 13*, 5315-5322.

Gordenin, D. A., Kunkel, T. A., & Resnick, M. A. (1997) *Nat. Genet. 16*, 116-118.

Harding, A. E. (1981) *Brain 104*, 589-620.

Heus, H. A., & Pardi, A. (1991) *Science 253*, 191-194.

Hiom, K., & Gellert, M. (1997) *Cell 88*, 65-72.

Huertas, D., Bellsolell, E. H., Casanovas, J. M., Coll, M., & Azorin, F. (1993) *EMBO J. 12*, 4029-4038.

Hunter, W. N., Brown, T., & Kennard, O. (1986) *J. Biomol. Struct. Dnyam. 4*, 173-191.

Johnston, B. H. (1992) *Methods in Enzymology 212*, 180-194.

Johnston, B. H., & Rich, A. (1985) *Cell 42*, 713-724.

Kang, S., Jaworski, A., Ohshima, K., & Wells, R. D. (1995) *Nature Genetics 10*, 213-218.

Kang, S., Ohshima, K., Jaworski, A., & Wells, R. D. (1996) *J. Mol. Biol. 258*, 543-547.

Kennard, O., & Hunter, W. N. (1989) *Quar. Rev. Biophys. 22*, 327-379.

Krasilnikov, A. S., Panyutin, I. G., Samadashwily, G. M., Cox, R., Lazurkin, Y. S., & Mirkin, S. M. (1997) *Nucleic Acids Res. 25,* 1339-1346.

Krol, A., & Carbon, P. (1989) *Methods in Enzymology, 180,* 212-227.

Lieber, M. R. (1997) *Bioessays 19,* 233-240.

Lohman, T. M., & Bjornson, K. P. (1996) *Ann. Rev. Biochem. 65,* 169-214.

Mariappan, S. V., Chen, X., Moyzis, R. K., Bradbury, E. M., & Gupta, G. (1997) *Biophysical J. 72,* A135.

Mariappan, S. V. S., Catasti, P., Chen, X., Ratcliff, R., Moyzis, R. K., Bradbury, E. M., & Gupta, G. (1996a) *Nucleic Acids Res. 24,* 784-792.

Mariappan, S. V. S., Garcia, A., & Gupta, G. (1996b) *Nucleic Acids Res. 24,* 775-783.

Maxam, A. M., & Gilbert, W. (1980) *Meth. Enzymol. 65,* 499-560.

McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., & Oettinger, M. A. (1995) *Cell 83,* 387-95.

McCarthy, J. G., Williams, L. D., & Rich, A. (1990) *Biochemistry 29,* 6071-6081.

McMurray, C. T., Gacy, A. M., Goellner, G. G., Spiro, C., Dyer, R., Yao, J., & Mikesell, M. (1997) *Cambridge symposia 5,* 22.

Mitas, M. (1997) *Nucleic Acids Res. 25,* 2245-2253.

Mitas, M., Yu, A., Dill, J., & Haworth, I. S. (1995a) *Biochemistry 34*, 12803-12811.

Mitas, M., Yu, A., Dill, J., Kamp, T. J., Chambers, E. J., & Haworth, I. S. (1995b) *Nucleic Acids Res. 23*, 1050-1059.

Mukerji, I., Shiber, M. C., Fresco, J. R., & Spiro, T. G. (1996) *Nucleic Acids Res. 24*, 5013-5020.

Mukerji, I., Shiber, M. C., Spiro, T. G., & Fresco, J. R. (1995) *Biochemistry 34*, 14300-14303.

Ohshima, K., Kang, S., Larson, J. E., & Wells, R. D. (1996a) *J. Biol. Chem. 271*, 16773-16783.

Ohshima, K., Kang, S., & Wells, R. D. (1996b) *J. Biol. Chem. 271*, 1853-1856.

Orita, M., Nishikawa, F., Kohno, T., Senda, T., Mitsui, Y., Endo, Y., Taira, K., & Nishikawa, S. (1996) *Nucleic Acids Res. 24*, 611-618.

Pearson, C. E., & Sinden, R. R. (1996) *Biochemistry 35*, 5041-53.

Pieretti. M., F. Zhang, Fu, Y. H, Warren, S.T., Oostra, B,A., Caskey, C.T., and Nelson, D.L. (1991) *Cell 66*, 817-822.

Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498-504.

Rao, B. S. (1994) *Gene 140*, 233-237.

Riggins, G. J., Lokey, L. K., Chastain, J. L., Leiner, H. A., Sherman, S. L., Wilkinson, K. D., & Warren, S. T.(1992) *Nat. Genet. 2*, 186-191.

Rippe, K., Fritsch, V., Westhof, E., & Jovin, T. M. (1992) *EMBO J. 11*, 3777-3786.

Robinson, H., van Boom, J. H., & Wang, A. H.-J. (1994) *J. Am. Chem. Soc. 116*, 1565-1566.

Rubin, C. M., & Schmid, C. W. (1980) *Nucleic Acids Res. 8*, 4613-4619.

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular cloning*, 2nd ed., Cold Spring Harbor Press, Plainview, NY.

Shiber, M. C., Braswell, E. H., Klump, H., & Fresco, J. R. (1996) *Nucleic Acids Res. 24*, 5004-5012.

Shiber, M. C., Lavelle, L., Fossella, J. A., & Fresco, J. R. (1995) *Biochemistry 34*, 14293-14299.

Singer, B., & Grunberger, D. (1983) *Molecular biology of mutagens and carcinogens*, Plenum Press, New York.

Smith, G. K., Jie, J., Fox, G. E., & Gao, X. (1995) *Nucleic Acids Res. 23*, 4303-4311.

Sutherland, G. R., & Richards, R. I. (1995) *Proc. Natl. Acad. Sci. USA 92*, 3636-3641.

Timchenko, L. T., & Caskey, C. T. (1996) *FASEB J. 10*, 1589-1597.

Usdin, K., & Woodford, K. J. (1995) *Nucleic Acids Res. 23*, 4202-4209.

Vincent, A., Heitz, D., Petit, C., Kretz, C., Oberle, I., & Mandel, J. L. (1991) Nature 369, 624-626.

Warren, S. T. (1996) *Science 271*, 1374-1375.

Wartell, R. M., Hosseini, S. H., & Moran, C. P. J. (1990) *Nucleic Acids Res. 18*, 2699-2705.

Wohlrab, F. (1992) *Meth. Enzymol. 212B*, 294-301.

Yoshizawa, S., Kawai, G., Watanabe, K., Miura, K., & Hirao, I. (1997) *Biochemistry 36*, 4761-4767.

Yu, A., Barron, M. D., Romero, R. M., Christy, M., Gold, B., Dai, J., Gray, D. M., Haworth, I. S., & Mitas, M. (1997a) *Biochemistry 36*, 3687-3699.

Yu, A., Dill, J., & Mitas, M. (1995a) *Nucleic Acids Research 23*, 4055-4057.

Yu, A., Dill, J., Wirth, S. S., Huang, G., Lee, V. H., Haworth, I. S., & Mitas, M. (1995b) *Nucleic Acids Research 23*, 2706-2714.

Table

Table 1. Sequences of oligonucleotides described in text

A. Oligonucleotides derived from plasmid

Name	Sequence	Number of nucleotides
oligo ***1***	AGCTTTGATCA(GAA)$_{15}$GGATC	61
oligo ***2***	AGCTTTGATCA(GAA)(GAT)(GAA)$_{14}$GGATC	64
oligo ***3***	AACC(GAA)$_{15}$GGATC	54
oligo ***1'***	GATCC(TTC)$_{15}$TGATCAAAGCT	61
oligo ***2'***	GATCC(TTC)$_{14}$(ATC)(TTC)TGATCAAAGCT	64
oligo ***3'***	GATCC(TTC)$_{15}$GGTT	54

B. Unlabeled synthetic oligonucleotides

oligo (***1***)	AGCTTTGATCA(GAA)$_{15}$G	57
oligo (***1'***)	GATCC(TTC)$_{15}$TGATCAA	57
oligo (***3***)	AGCTTGTTAACC(GAA)$_{15}$G	58
oligo (***3'***)	GATCC(TTC)$_{15}$GGTTAACA	58

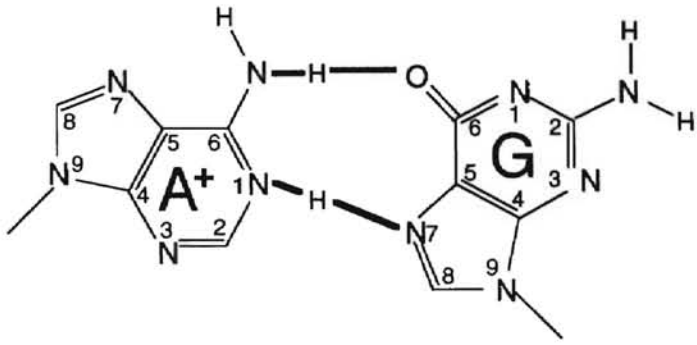
Sequences of the oligonucleotides are written 5'→ 3'.

Underlined nucleotides in oligonucleotides derived from plasmids are those added by Klenow enzyme. Nucleotides doubly underlined contain a 5' ^{32}P label.

Figure 1. Three types of A•G base pairs that have been observed in crystal structures of duplex DNA (Kennard & Hunter, 1989).

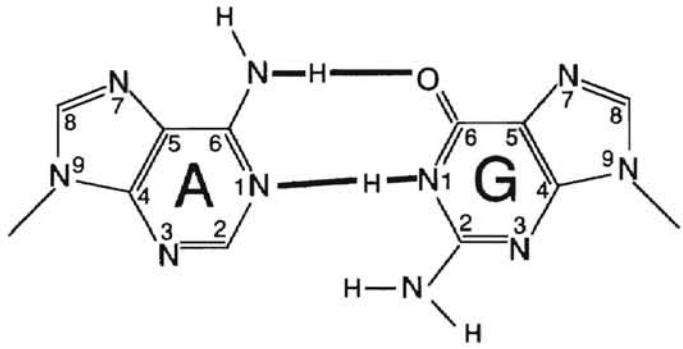
Each of these A•G base pairs [A^{syn}•G^{anti} (Brown et al., 1986; Hunter et al., 1986), A^{anti}•G^{anti} (Brown et al., 1989), and A^{+anti}•G^{syn} (Prive et al., 1987)] contains two H-bonds, one of which is adenine N6-guanine O6.

A



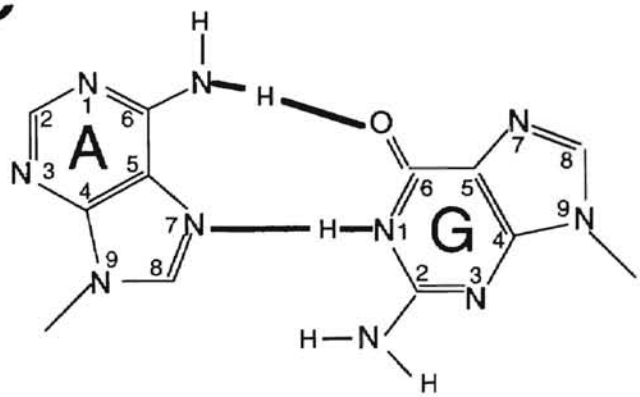
A^{+}(*anti*) • G(*syn*)

B



A(*anti*) • G(*anti*)

C



A(*syn*) • G(*anti*)

Figure 2. Chemical modification and P1 nuclease analysis of oligo ***2'***.

Oligo ***2'***, the complete sequence of which is shown to the left of the figure, was isolated from plasmid pGAA2 and incubated with 4.3 M hydroxylamine (left lane) and 0.1 units of P1 nuclease (right lane) for 25 min as described in materials and methods. Roman numerals correspond to triplet repeats. Arabic numerals indicate the position of a nucleotide with respect to the 5'-end. Samples were applied to a DNA sequencing gel containing 8 M urea and 20% polyacrylamide. Bands in the control lane correspond to depurination products. Arrows on right indicate positions of P1 nuclease cleavage sites of selected sequences.

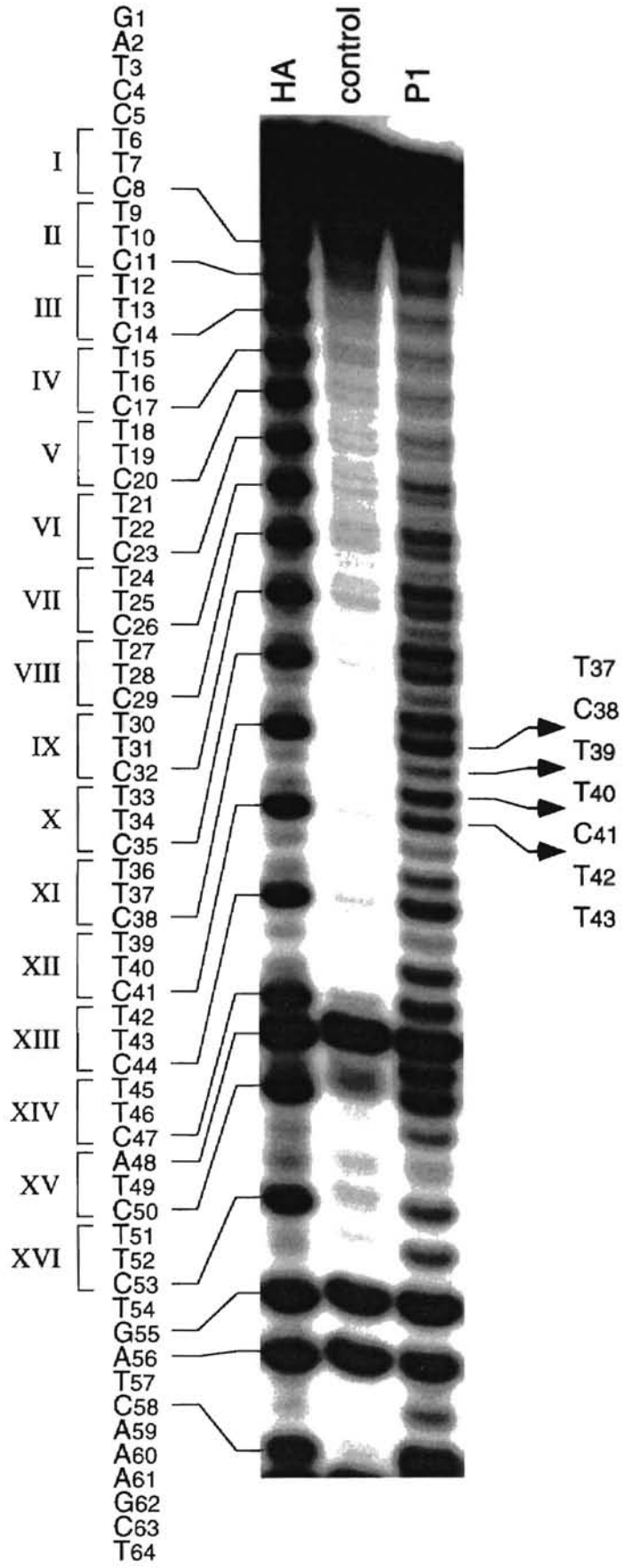

HA
control
P1
G1
A2
T3
C4
C5
I
T6
T7
C8
II
T9
T10
C11
III
T12
T13
C14
IV
T15
T16
C17
V
T18
T19
C20
VI
T21
T22
C23
VII
T24
T25
C26
VIII
T27
T28
C29
IX
T30
T31
C32
X
T33
T34
C35
XI
T36
T37
C38
XII
T39
T40
C41
XIII
T42
T43
C44
XIV
T45
T46
C47
XV
A48
T49
C50
XVI
T51
T52
C53
T54
G55
A56
T57
C58
A59
A60
A61
G62
C63
T64
T37
C38
T39
T40
C41
T42
T43

Figure 3. Potential structures adopted by oligo ***3***.

Five potential structures of oligo ***3*** are shown. Structures ***A***, ***B***, and ***D*** are various hairpin alignments containing A•G (thick lines), A•A (thin lines), or G•G (open ovals) base-pairs. ***C*** is a linear structure stabilized primarily by stacking interactions. The triplet repeat region in structure ***E*** is arranged in a helical array.

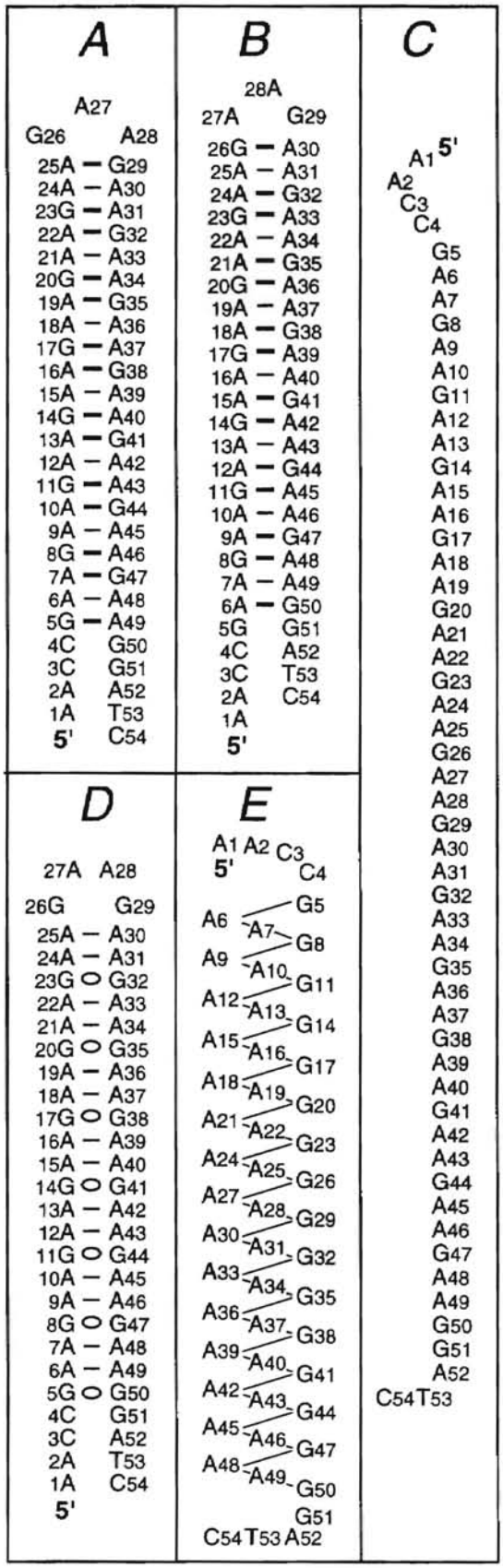

A
A27
G26 A28
25A — G29
24A — A30
23G — A31
22A — G32
21A — A33
20G — A34
19A — G35
18A — A36
17G — A37
16A — G38
15A — A39
14G — A40
13A — G41
12A — A42
11G — A43
10A — G44
9A — A45
8G — A46
7A — G47
6A — A48
5G — A49
4C G50
3C G51
2A A52
1A T53
5' C54
B
28A
27A G29
26G — A30
25A — A31
24A — G32
23G — A33
22A — A34
21A — G35
20G — A36
19A — A37
18A — G38
17G — A39
16A — A40
15A — G41
14G — A42
13A — A43
12A — G44
11G — A45
10A — A46
9A — G47
8G — A48
7A — A49
6A — G50
5G G51
4C A52
3C T53
2A C54
1A
5'
C
A1 5'
A2
C3
C4
G5
A6
A7
G8
A9
A10
G11
A12
A13
G14
A15
A16
G17
A18
A19
G20
A21
A22
G23
A24
A25
G26
A27
A28
G29
A30
A31
G32
A33
A34
G35
A36
A37
G38
A39
A40
G41
A42
A43
G44
A45
A46
G47
A48
A49
G50
G51
A52
C54 T53
D
27A A28
26G G29
25A — A30
24A — A31
23G O G32
22A — A33
21A — A34
20G O G35
19A — A36
18A — A37
17G O G38
16A — A39
15A — A40
14G O G41
13A — A42
12A — A43
11G O G44
10A — A45
9A — A46
8G O G47
7A — A48
6A — A49
5G O G50
4C G51
3C A52
2A T53
1A C54
5'
E
A1 A2 C3
5' C4
A6 G5
A7
A9 G8
A10
A12 G11
A13
A15 G14
A16
A18 G17
A19
A21 G20
A22
A24 G23
A25
A27 G26
A28
A30 G29
A31
A33 G32
A34
A36 G35
A37
A39 G38
A40
A42 G41
A43
A45 G44
A46
A48 G47
A49
G50
G51
C54 T53 A52

Figure 4. Diethylpyrocarbonate modification of oligo ***3***.

Oligo ***3***, the complete sequence of which is shown on the left, was isolated from plasmid pGAA3 and incubated with 0.46 M diethylpyrocarbonate at the indicated temperatures as described in materials and methods. Arrows point to adenines that were extensively modified by DEPC at 5 °C. A hairpin structure that is consistent with the modification data is shown to the left. Lines between bases indicate formation of presumed H-bonds. Roman numerals correspond to triplet repeats. Arabic numerals indicate the position of a nucleotide with respect to the 5'-end. Arrows indicate adenine residues preferentially modified at 5 °C.

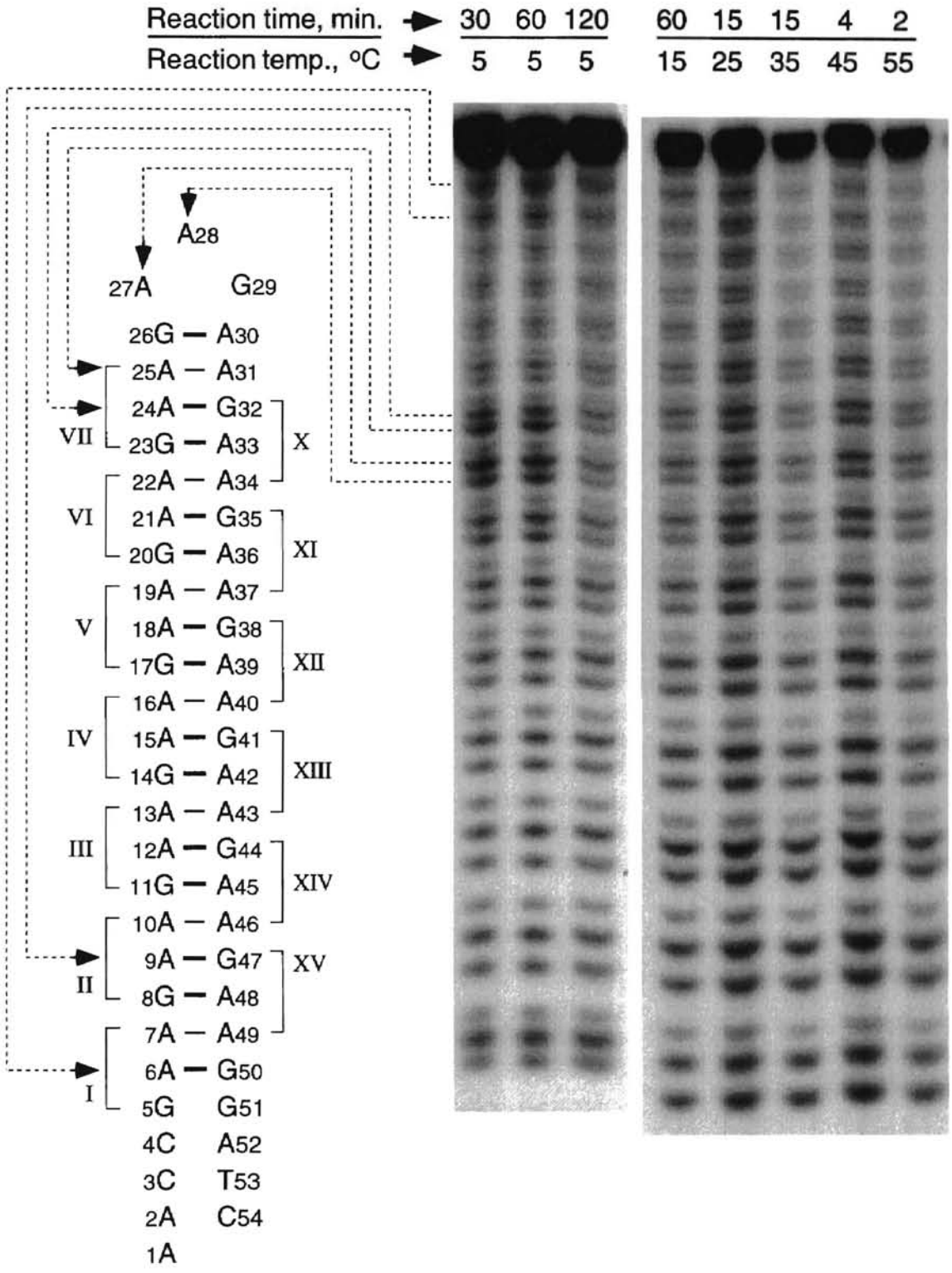

Reaction time, min. → 30 60 120 60 15 15 4 2
Reaction temp., °C → 5 5 5 15 25 35 45 55
A28
27A G29
26G — A30
25A — A31
24A — G32
23G — A33
22A — A34
21A — G35
20G — A36
19A — A37
18A — G38
17G — A39
16A — A40
15A — G41
14G — A42
13A — A43
12A — G44
11G — A45
10A — A46
9A — G47
8G — A48
7A — A49
6A — G50
5G G51
4C A52
3C T53
2A C54
1A
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV

Figure 5. P1 nuclease digestion analysis of oligo ***3*** at 50 mM Na^+ and various temperature.

Oligo ***3***, the complete sequence of which is shown on the left, was isolated from plasmid pGAA3 and incubated with P1 nuclease. Samples were applied to a DNA sequencing gel containing 8 M urea and 20% polyacrylamide. P1 nuclease digestion in 50 mM Na^+. Reactions were performed at the indicated temperature in 50 mM NaCl, pH 7.5. From left to right, the amount of P1 nuclease (and time of incubation) added in each reaction were, $9x10^{-2}$ U (10 min), $2.2x10^{-2}$ U (5 min), $2.8x10^{-3}$ U (5 min), and $2.8x10^{-3}$ U (5 min). Dimethyl sulfate (21 mM) modification was performed at 25 °C in 50 mM NaCl as described in materials and methods. A hairpin structure that is consistent with the P1 nuclease data at 5 °C is shown to the left. Arrows indicate phosphodiester bonds cleaved by P1 nuclease.

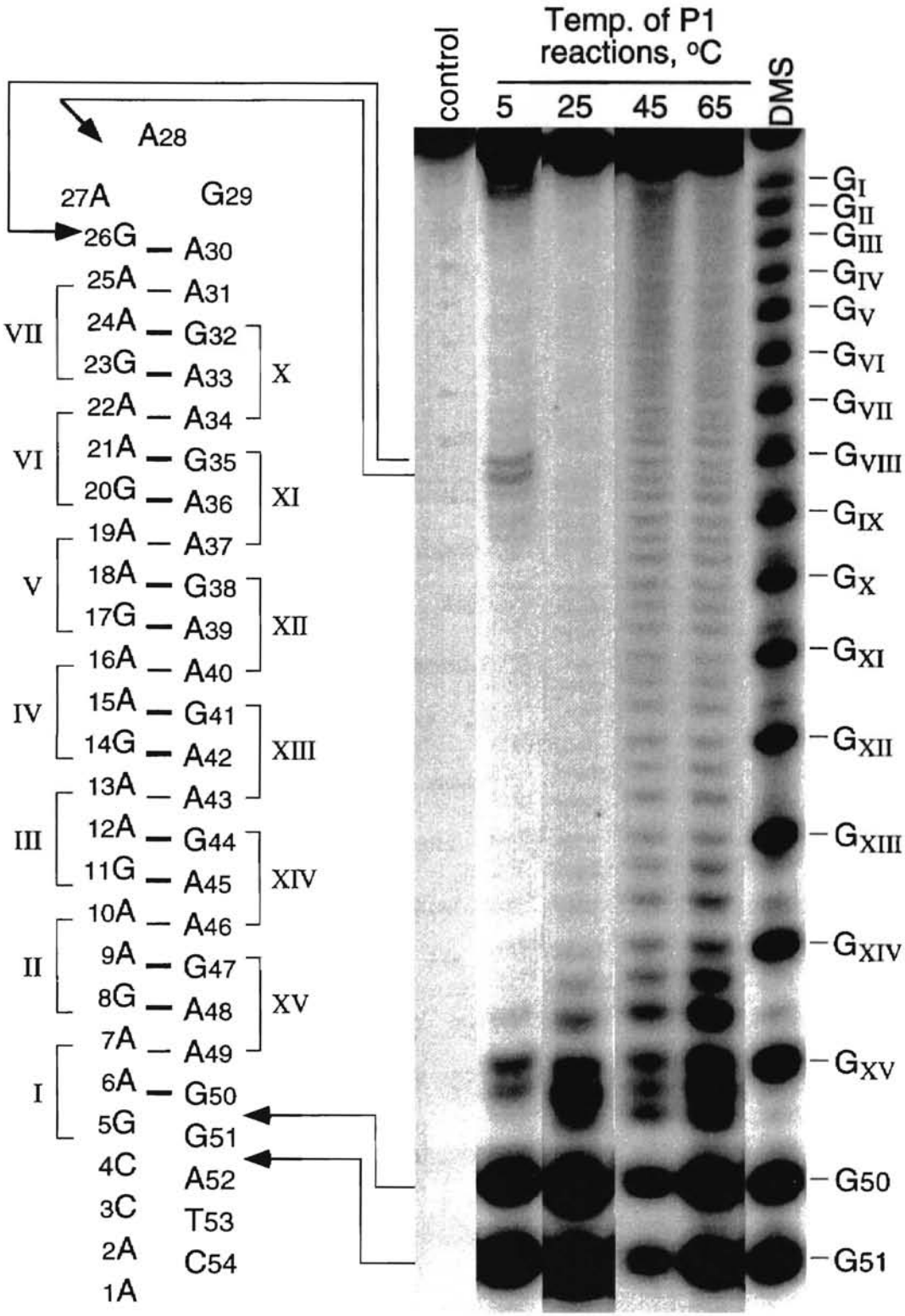

control
Temp. of P1 reactions, °C
5
25
45
65
DMS
A28
27A
G29
26G
A30
25A
A31
VII
24A
G32
23G
A33
X
22A
A34
VI
21A
G35
20G
A36
XI
19A
A37
V
18A
G38
17G
A39
XII
16A
A40
IV
15A
G41
14G
A42
XIII
13A
A43
III
12A
G44
11G
A45
XIV
10A
A46
II
9A
G47
8G
A48
XV
7A
A49
I
6A
G50
5G
G51
4C
A52
3C
T53
2A
C54
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 6. P1 nuclease digestion analysis of oligo ***3*** at 50 mM Na^+, 15 °C.

Oligo ***3***, the complete sequence of which is shown on the left, was isolated from plasmid pGAA3 and incubated with P1 nuclease. Samples were applied to a DNA sequencing gel containing 8 M urea and 20% polyacrylamide. P1 nuclease digestion in 50 mM Na^+. Reactions were performed at 15 °C, 50 mM NaCl, pH 7.5. From left to right, the amount of P1 nuclease added in each reaction were, 1.1×10^{-2} U, 4.4×10^{-2} U and 8.8×10^{-2} U. The time of incubation was 6 min. Dimethyl sulfate (21 mM) modification was performed at 5 °C in 50 mM NaCl. A hairpin structure that is consistent with the P1 nuclease data at 15 °C is shown to the left. Arrows indicate phosphodiester bonds cleaved by P1 nuclease.

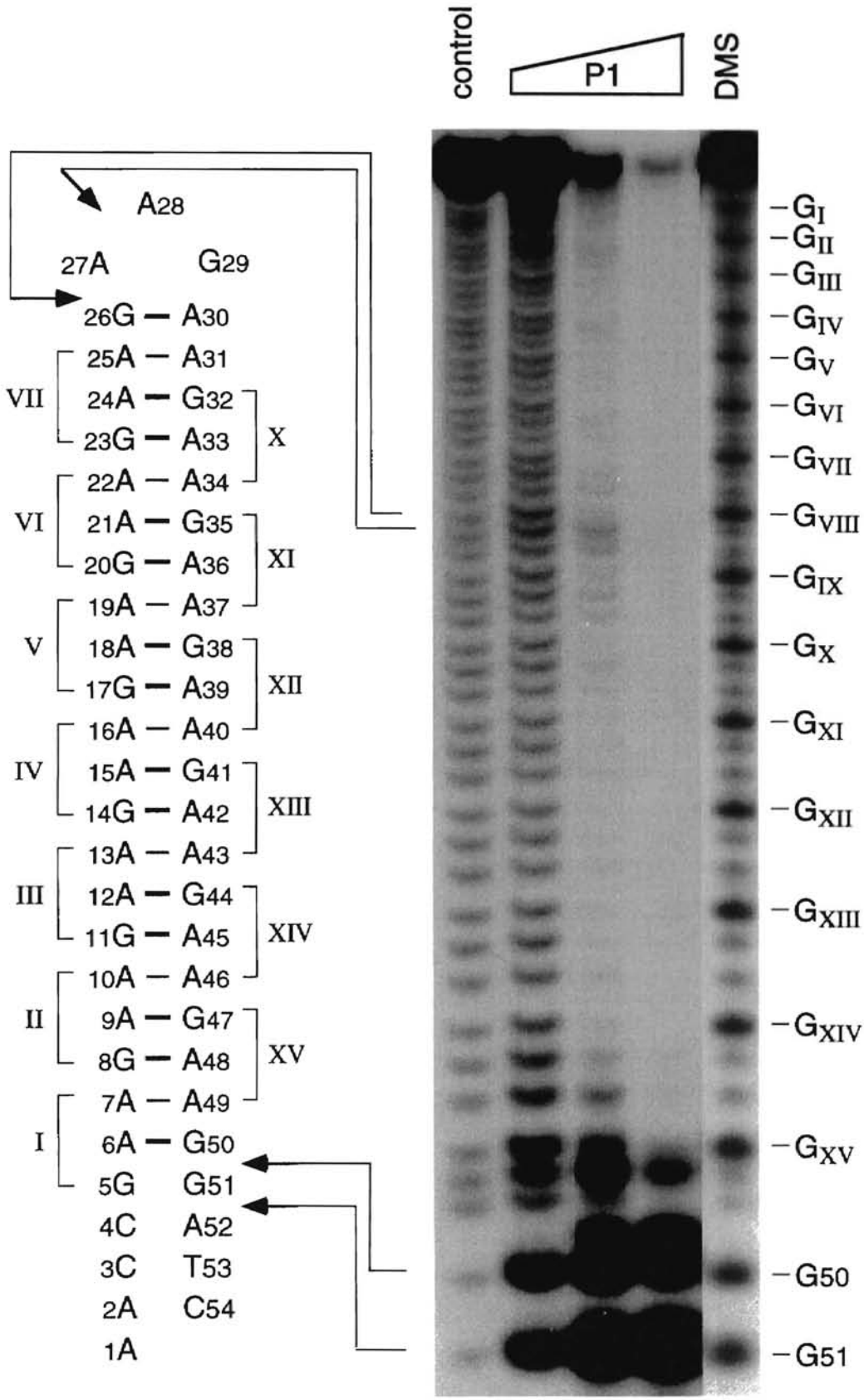

control
P1
DMS
A28
27A
G29
26G — A30
25A — A31
VII
24A — G32
23G — A33
X
22A — A34
VI
21A — G35
20G — A36
XI
19A — A37
V
18A — G38
17G — A39
XII
16A — A40
IV
15A — G41
14G — A42
XIII
13A — A43
III
12A — G44
11G — A45
XIV
10A — A46
II
9A — G47
8G — A48
XV
7A — A49
I
6A — G50
5G G51
4C A52
3C T53
2A C54
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 7. P1 nuclease digestion analysis of oligo ***3*** at 37 °C 50mM Na^+.

Oligo ***3***, the complete sequence of which is shown on the left, was isolated from plasmid pGAA3 and incubated with P1 nuclease. Samples were applied to a DNA sequencing gel containing 8 M urea and 20% polyacrylamide. Reactions were performed at pH 7.5 ,37 °C. The amount of P1 nuclease added from left to right were, $1.1x10^{-2}$ U, $2.2x10^{-2}$ U and $5.0x10^{-2}$ U. The control lane contained no P1 nuclease. The time of incubation was 6 min. Dimethyl sulfate (21 mM) modification was performed at 25 °C in 50 mM NaCl. Arrows indicate phosphodiester bonds cleaved by P1 nuclease.

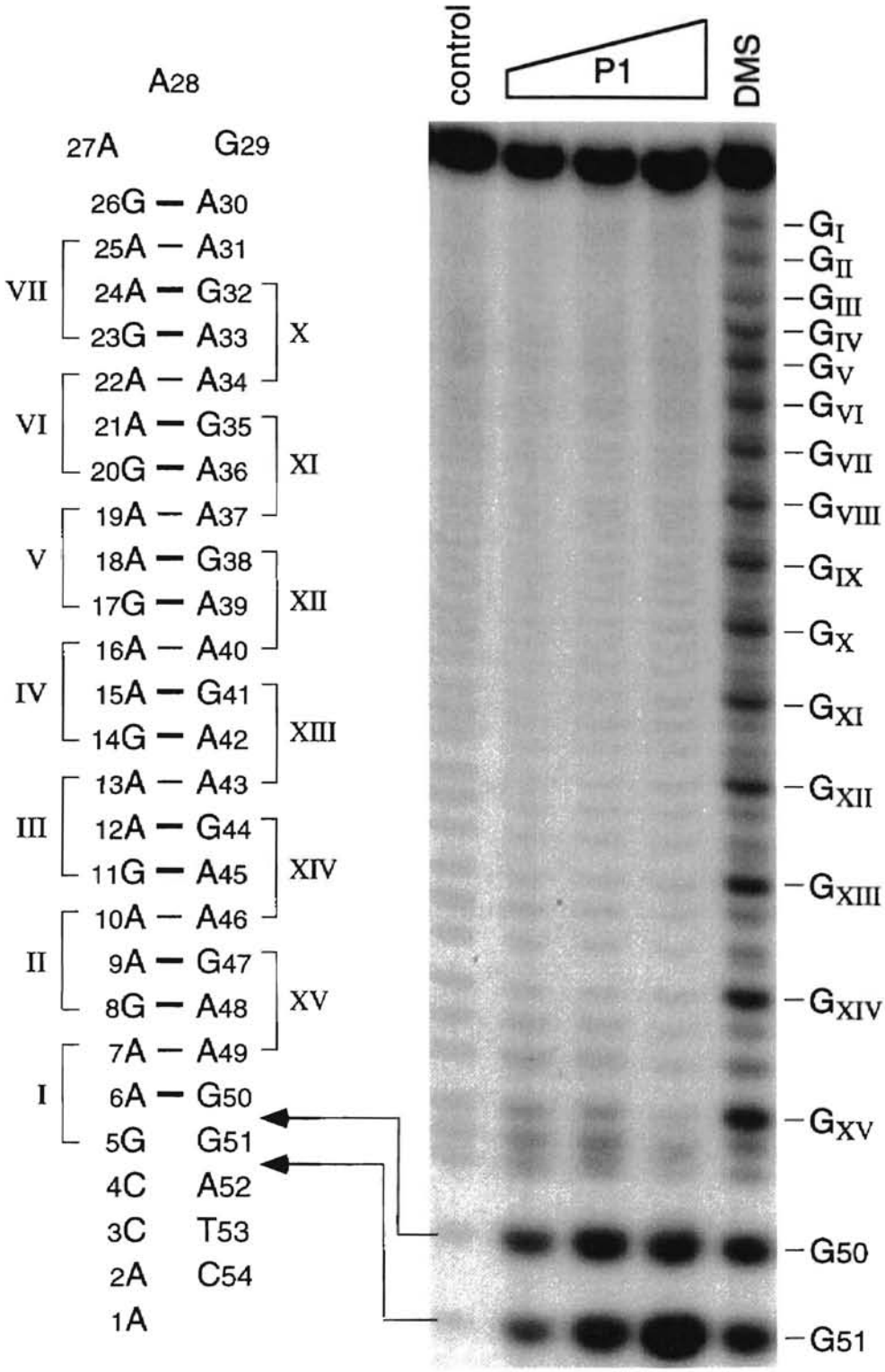

control
P1
DMS
A28
27A G29
26G — A30
25A — A31
24A — G32
23G — A33
22A — A34
21A — G35
20G — A36
19A — A37
18A — G38
17G — A39
16A — A40
15A — G41
14G — A42
13A — A43
12A — G44
11G — A45
10A — A46
9A — G47
8G — A48
7A — A49
6A — G50
5G G51
4C A52
3C T53
2A C54
1A
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 8. P1 nuclease digestion analysis of oligo ***3*** at 50 mM Na^+ ,5 °C and different pH.

P1 nuclease digestion in 50 mM Na^+. Reactions were performed at the indicated pH in 50 mM NaCl, 5 °C. The amount of P1 nuclease (and time of incubation) added in each reaction were, $2.2x10^{-2}$ U (15 min), $4.4x10^{-2}$ U (15 min) and $8.8x10^{-3}$ U (7.5 min). Dimethyl sulfate (21 mM) modification was performed at 5 °C in 50 mM NaCl as described in materials and methods. A hairpin structure that is consistent with the P1 nuclease data is shown to the left. Arrows indicate phosphodiester bonds cleaved by P1 nuclease.

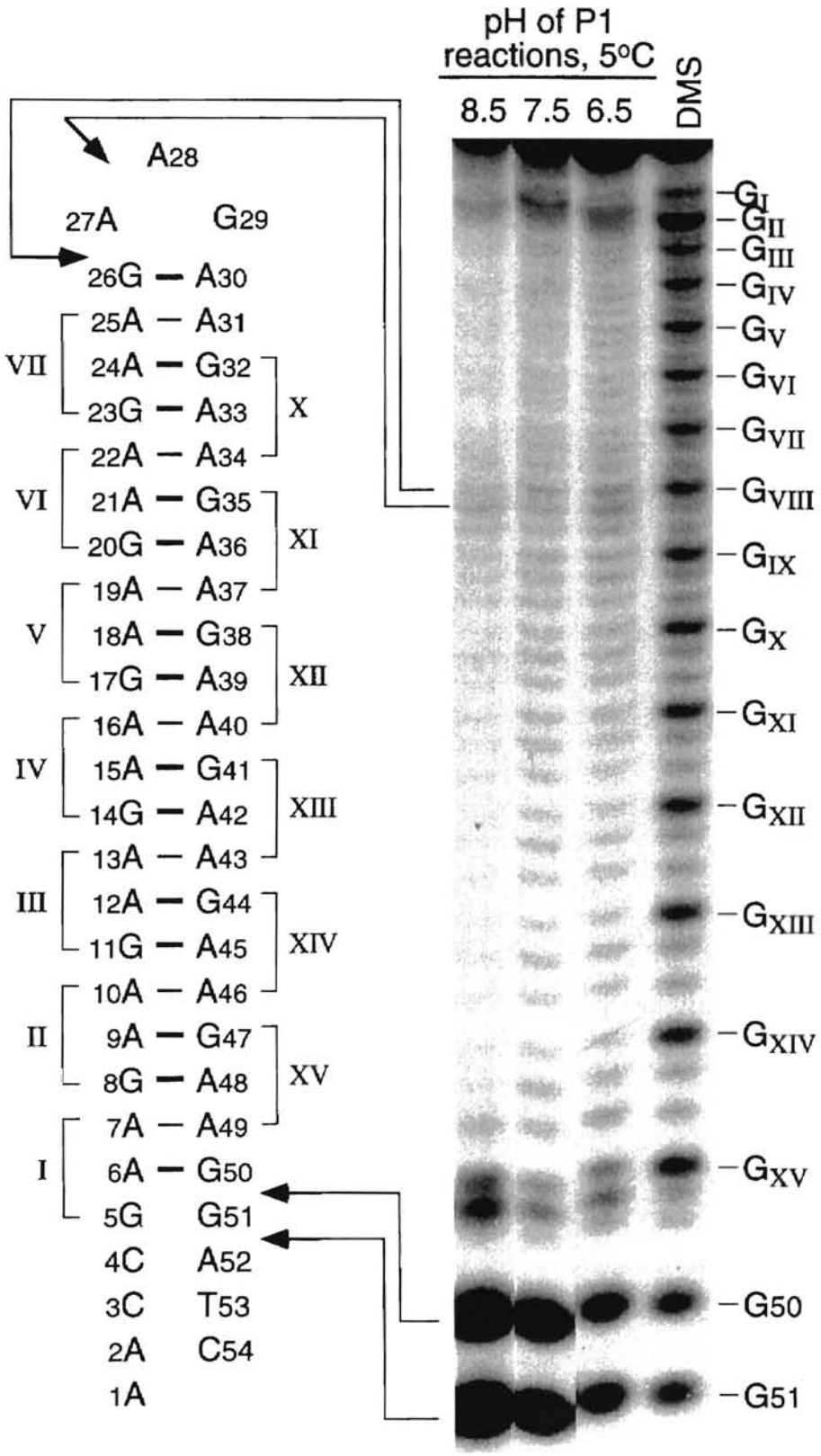

pH of P1 reactions, 5°C
8.5 7.5 6.5
DMS
A28
27A G29
26G — A30
25A — A31
VII 24A — G32
23G — A33 X
22A — A34
VI 21A — G35
20G — A36 XI
19A — A37
V 18A — G38
17G — A39 XII
16A — A40
IV 15A — G41
14G — A42 XIII
13A — A43
III 12A — G44
11G — A45 XIV
10A — A46
II 9A — G47
8G — A48 XV
7A — A49
I 6A — G50
5G G51
4C A52
3C T53
2A C54
1A
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 9. P1 nuclease digestion analysis of oligo ***3*** at 150 mM Na^+ ,5 °C.

P1 nuclease digestion in 150 mM Na^+. Reactions were performed at the indicated pH in 150 mM NaCl, pH 7.5, 5 °C. The amount of P1 nuclease added was 0.1 U. The time of incubation was 20 min. Dimethyl sulfate (21 mM) modification was performed at 5 °C in 50 mM NaCl as described in materials and methods. A hairpin structure that is consistent with the P1 nuclease data is shown to the left. Arrows indicate phosphodiester bonds cleaved by P1 nuclease.

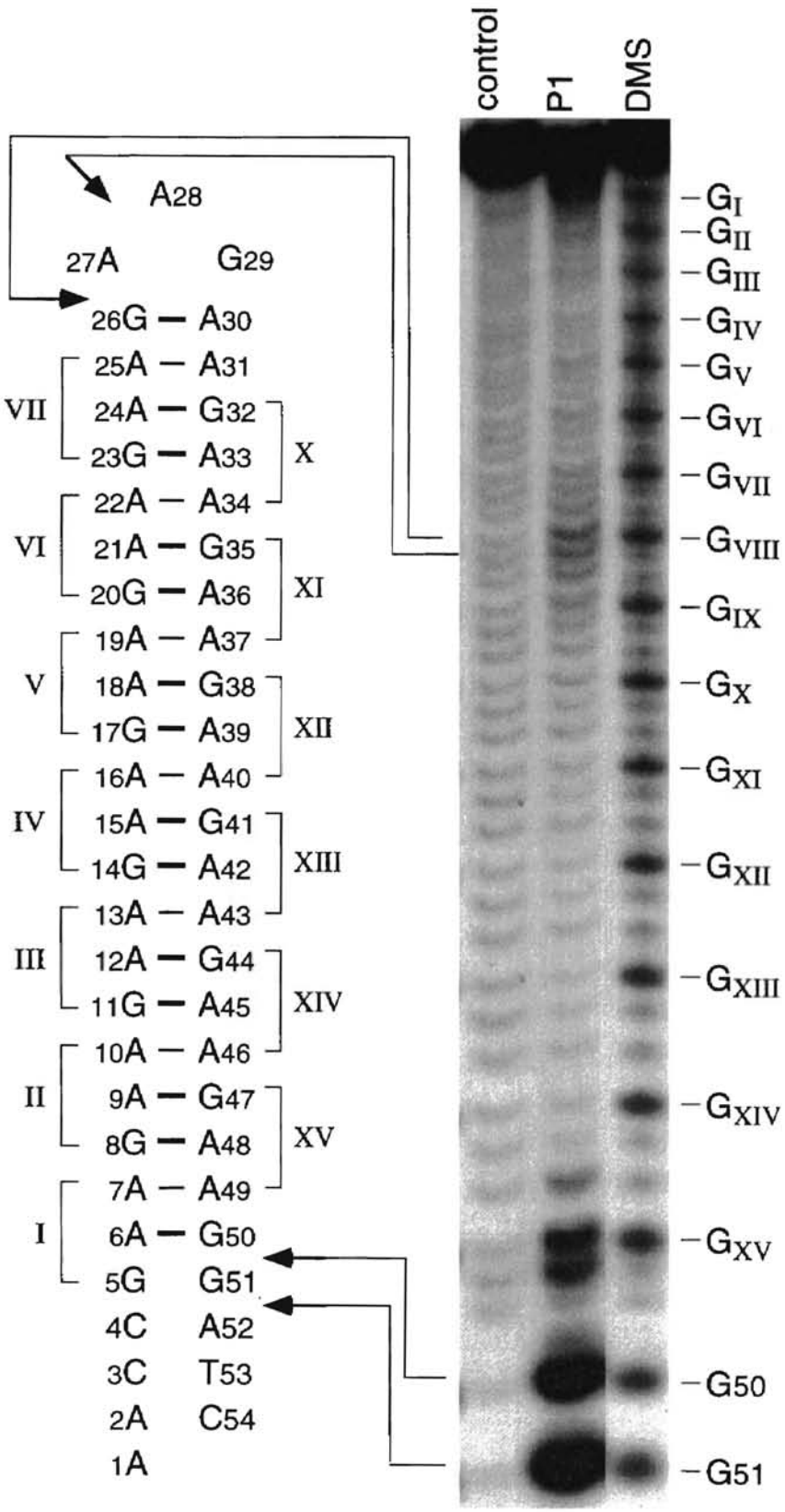

control
P1
DMS
A28
27A
G29
26G — A30
25A — A31
24A — G32
23G — A33
22A — A34
21A — G35
20G — A36
19A — A37
18A — G38
17G — A39
16A — A40
15A — G41
14G — A42
13A — A43
12A — G44
11G — A45
10A — A46
9A — G47
8G — A48
7A — A49
6A — G50
5G G51
4C A52
3C T53
2A C54
1A
VII
VI
V
IV
III
II
I
X
XI
XII
XIII
XIV
XV
GI
GII
GIII
GIV
GV
GVI
GVII
GVIII
GIX
GX
GXI
GXII
GXIII
GXIV
GXV
G50
G51

Figure 10. P1 nuclease digestion analysis of oligo ***2*** at 37 °C and various amounts of Na^+ concentrations.

Oligo ***2***, the complete sequence of which is shown on the left, was isolated from plasmid pGAA2 and incubated with various amounts of P1 nuclease at 37 °C in various amount of NaCl, pH 7.5. In each set of reaction, the amounts of P1 nuclease added from left to right were 0.013, 0.025, 0.05, and 0.10 units. The control lane contained no P1 nuclease. The positions of the guanine nucleotides in the control lane, which arose from minor depurination, are shown. Samples were applied to a DNA sequencing gel containing 8 M urea and 20% polyacrylamide. Roman numerals correspond to triplet repeats. Arabic numerals indicate the position of a nucleotide with respect to the 5'-end. Arrows indicate phosphodiester bonds cleaved with P1 nuclease.

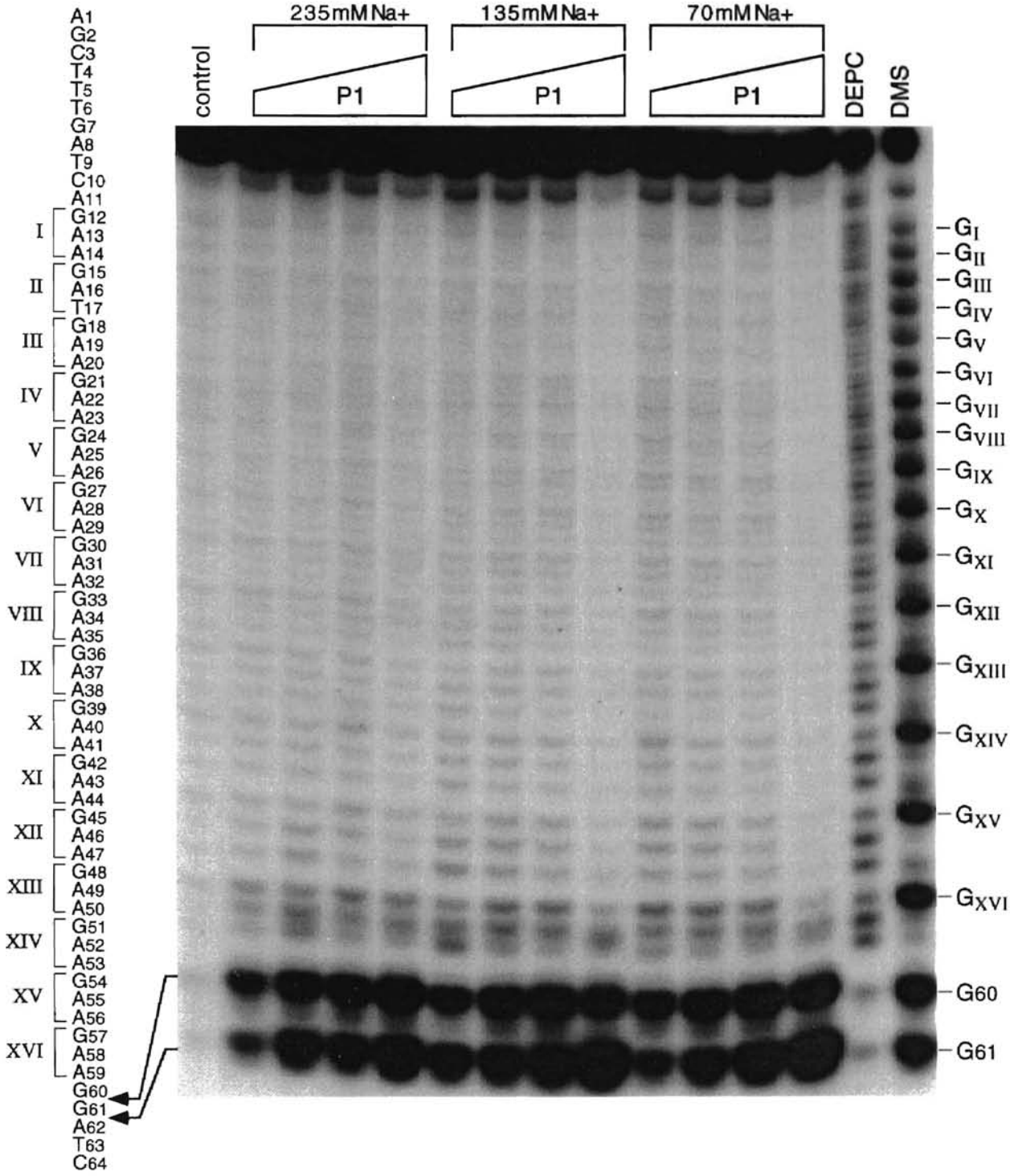
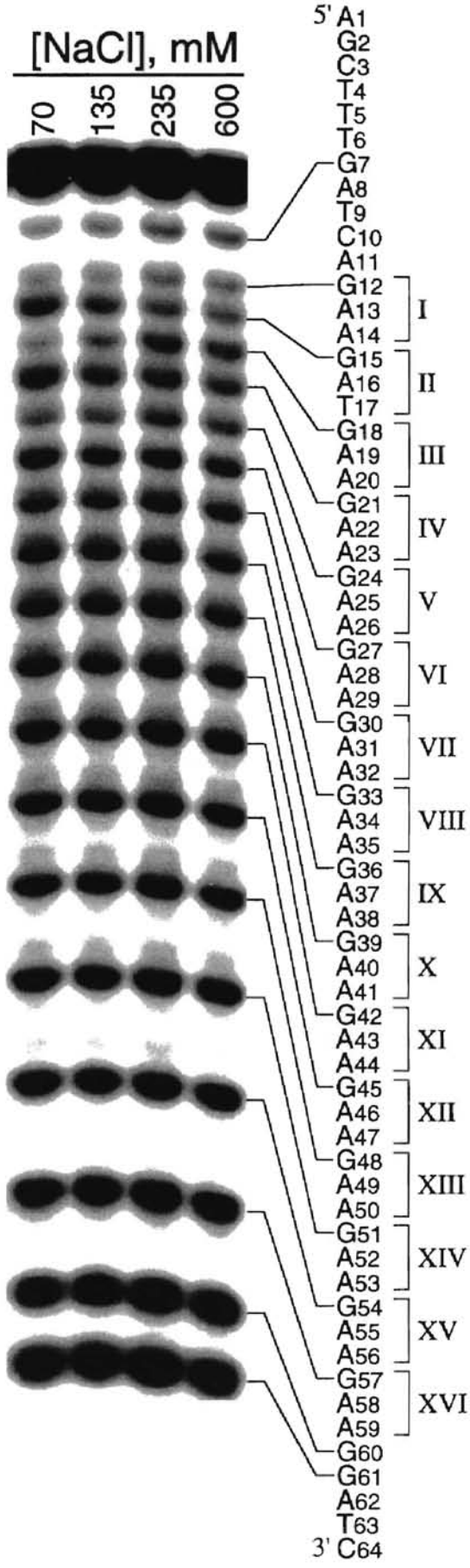

235mM Na+
135mM Na+
70mM Na+
control
P1
P1
P1
DEPC
DMS
A1 G2 C3 T4 T5 T6 G7 A8 T9 C10 A11
I G12 A13 A14
II G15 A16 T17
III G18 A19 A20
IV G21 A22 A23
V G24 A25 A26
VI G27 A28 A29
VII G30 A31 A32
VIII G33 A34 A35
IX G36 A37 A38
X G39 A40 A41
XI G42 A43 A44
XII G45 A46 A47
XIII G48 A49 A50
XIV G51 A52 A53
XV G54 A55 A56
XVI G57 A58 A59
G60 G61 A62 T63 C64
GI GII GIII GIV GV GVI GVII GVIII GIX GX GXI GXII GXIII GXIV GXV GXVI
G60
G61

Figure 11. Dimethyl sulfate modification of oligo ***2*** .

Oligo ***2***, the complete sequence of which is shown on the right, was isolated from plasmid pGAA2 and incubated with 21 mM dimethyl sulfate at room temperature, pH 7.5, at the indicated salt concentration as described in materials and methods. Lines indicate positions of guanine bases.


[NaCl], mM
70
135
235
600
5' A1
G2
C3
T4
T5
T6
G7
A8
T9
C10
A11
G12
A13
A14
G15
A16
T17
G18
A19
A20
G21
A22
A23
G24
A25
A26
G27
A28
A29
G30
A31
A32
G33
A34
A35
G36
A37
A38
G39
A40
A41
G42
A43
A44
G45
A46
A47
G48
A49
A50
G51
A52
A53
G54
A55
A56
G57
A58
A59
G60
G61
A62
T63
3' C64
I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
XIV
XV
XVI

VITA

Iang-Shan Suen

Candidate for the Degree of

Master of Sciences

Thesis: THE FRIEDREICH'S ATAXIA TRIPLET REPEAT SEQUENCE $d(GAA)_n$ FORMS A STABLE UNIMOLECULAR STRUCTURE: EVIDENCE FOR A HAIRPIN CONTAINING $A^{anti}\cdot G^{anti}$ BASE PAIRS

Major Field: Biochemistry and Molecular Biology

Biographical:

Personal Data: Born in Taipei, Taiwan, R. O. China, February 6th, 1969, the son of Tzeng-Rong Suen and An-Hua Ping.

Education: Received Bachelor of Science Degree from Tunghai University, Taichung, Taiwan, R. O. China in July, 1992; completed requirements for Master of Science Degree at Oklahoma state University, U. S. A. in August, 1997.

Abstracts Presented at Meetings: The Adenylates of Friedreich's Ataxia d(GAA) Repeats Are Nuclease -Resistance and Predominantly Unstacked: Evidence For a New Helical Single-Stranded DNA Structure. (1997) *J. Bioml. Struct. Dyn. 14*, 792. Friedreich's Ataxia d(GAA) Repeats Form a Heat-Stable Nuclease-Resistant Structure. (1997) *Biophysical J. 72*, A135.

Professional Experiences: Research Assistant, Institute of Life Sciences, National Tsing Hua University, Taiwan, October, 1994 to July 1995. Graduate Research Associate, Department of Biochemistry & Molecular Biology, Oklahoma State University, August, 1995 to August, 1997.

Professional Membership: Member of American Association for the Advancement of Science; Member of Biophysical Society.